DOCUMENT RESUME

ED 093 945 | TM 003 763

AUTHOR	Bejar, Isaac I.; Doyle, Kenneth O.
TITLE	Generalizability of Factor Structures Underlying Student Ratings of Instruction.
PUB DATE	Apr 74
NOTE	26p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	Factor Analysis; *Factor Structure; *Generalization; Individual Differences; Instructional Improvement; Questionnaires; Rating Scales; Reliability; Student Teacher Relationship; *Teacher Rating; Teaching Quality

ABSTRACT

The generalizability of factor structures of student ratings of instruction based on instructors' individual differences was examined. The subjects were instructors from the humanities, social science, and mathematics divisions who had had their courses evaluated at least twice using the same evaluation questionnaire. The data from the three divisions were factor analyzed and the resulting factor structures compared using Kaiser's procedure. Only one factor ("defines student responsibilities") was found to be stable within the across divisions. The results were explained in terms of the distinction between within-class covariation and between-class covariation. (Author)

GENERALIZABILITY OF FACTOR STRUCTURES UNDERLYING STUDENT RATINGS OF INSTRUCTION

Isaac I. Bejar and Kenneth O. Doyle
University of Minnesota, Measurement Services Center

Paper presented at the annual meeting of the
American Educational Research Association.
Chicago, April 1974

ABSTRACT

Studied the generalizability of factors structures of student ratings of instruction based on instructors' individual differences. The subjects were instructors from the Humanities, Social Science and Science & Math division who had had their courses evaluated at least twice using the same evaluation questionnaire. The data from the three divisions were factor analyzed and the resulting factor structures compared using Kaiser's procedure. Only one factor ("defines student responsibilities") was found to be stable within and across divisions. The results were explained in terms of the distinction between within-class covariation and between-class covariation.

Generalizability of Factor Structures Underlying Student Ratings of Instruction[1]

Isaac I. Bejar and Kenneth O. Doyle Jr.[2]
University of Minnesota Measurement Services Center

Researchers over the years have devoted considerable effort to defining the underlying factor structure of students' ratings of instruction (e.g., Smalzreid and Remmers, 1943; Creager, 1950; Wherry, 1952; Crannell, 1953; Bendig, 1953, 1954; Coffman, 1954; Gibb, 1955; Isaacson, McKeachie, Milholland, Lin, Hofeller, Baerwaldt, and Zinn, 1964; Deshpande, Webb, and Marks, 1970; and Finkbeiner, Lathrop, and Schuerger, 1973). One outgrowth of these studies has been the relatively common practice of preparing concise data reports on the basis of a factor analysis so that an instructor, instead of receiving a printout of descriptive statistics for two dozen or more separate items, receives just four or five "factor scores" that summarize the ratings.

Some efforts to refine the concept of reliability have raised questions about the legitimacy of such procedures, however. Cattell (1954) and Cronbach, Gleser, Nada, and Rajaratnam (1972) talk about the "consistency" or "generalizability" of scores over various conditions such as people, occasions, and items. This line of thinking can be applied as well to factor structures as to other "scores." If factor structures do not vary over people or time or other conditions, these factor structures are generalizable or consistent. But when different students rate different instructors in different courses, it seems curious that the factor structures should be the same. Since factor structures of ratings are essentially statements of which instructor characteristics the students perceive to covary, one might expect these covariations to be instructor-specific. But if factor structures are instructor-specific -- i.e., are not consistent or generalizable -- then there would seem to be no rationale for computing factor scores on the basis of any one analysis. Under such circumstances, either factor scores cannot be used at all, or separate factor analyses need to be performed for every different condition.

At least two studies have directly addressed the question of generalizability of factor structures of student ratings. Isaacson et al (1964), with ratings on a 46-item instrument from more than 1200 students in 33 sections of a beginning course in psychology, used Kaiser's factor similarity technique (Kaiser, 1960; see also Kaiser, Hunka, Bianchini, 1971)[3] and found six factors to be consistent across sex of student and over two academic terms. Finkbeiner, Lathrop, and Schuerger (1973), with a 48-item instrument completed by almost 8,000 students in some 650 classes, employed Schneewind and Cattell's procedure (1971) and found a five-factor solution obtained on the main campus of a state university to be highly congruent with one obtained at academic centers.

From these studies it would appear that factor structures are generalizable over sex of student, academic term, and those characteristics that differentiate main-campus students from those at academic centers.

However, there is a question about the units of analysis employed in these studies. In both cases the ratings given to various instructors were merged, and the factor analyses were performed on a correlation matrix based on these pooled data matrices. The effect of this procedure is to confound two independent sources of covariation, the between-instructors and the pooled within-class covariation. Within-class covariation is obtained by intercorrelating item responses, while between-instructors covariation is gotten by intercorrelating item means. Because the calculation of means cancels out student individual differences (e.g., rater response tendencies; see Guilford, 1954, 278-89), between-instructors covariation would seem more descriptive of the instructors themselves.

The purpose of the present study, then, is to examine the generalizability of factor structures based on instructor individual differences.

METHOD

Rating Scale. The "Student Evaluation Form,Part I" includes 26 rating items on 5-point Always to Never scales. An earlier factor analysis (Doyle and Liu, 1972) using both alpha and principal factors solutions had found four factors tentatively named "Student-orientedness," "Organization," "Instructor Presence" (i.e., clear and enthusiastic), and "Intellectual Expansiveness" (i.e., broadminded).

Samples. Students in all courses in the Humanities, Social Science, and Science/Mathematics divisions of the Morris campus of the University of Minnesota during the Fall 1972 and Winter 1973 terms rated their instructors. These ratings were given in class toward the end of each term. The raters remained anonymous. From this pool of profiles, two samples were drawn, without replacement, in the following fashion. For the initial sample, an instructor's name was drawn at random. If he or she taught more than one course, freshman/sophomore courses were preferred to junior/senior ones. If two or more courses were still "eligible," one was selected at random. A repetition sample was chosen in similar fashion from the remaining profiles. Instructors who appeared in both samples were retained: 25 instructors in Humanities, 15 in Social Science, and 15 in Science/Mathematics.

Analyses. Item means were calculated for each class. Correlation matrices were computed on these means, one for each division in both the initial and the repetition sample. Factors were extracted by the method of principal axes, with iterations on the communalities. The largest off-diagonal values were taken as initial estimates of communality.[4] Retaining factors with eigenvalues greater than 1.0, four-, five-, and six-factor solutions were found. Because five-factor solutions were most frequent, the analyses were repeated with the number of factors to be extracted set at five. These factors were rotated to a varimax criterion. Kaiser et. al. method (1970) was used to compare these factors across and within divisions. Since no sampling distribution for the similarity coefficient exists, .70 was taken as the minimum indication of factor similarity. For the intradivisional comparisons, the factor structure from the repetition sample was rotated to similarity with the structure from the initial sample; for the interdivisional comparisons, the first sample of each pair in Table 1 below was the target on criterion structure.

RESULTS

Comparison across academic divisions. The five most similar pairs from each of the three interdivisional comparison are presented in Table 1, with factor similarity coefficients, indications of salient loadings, and fit. Salient loadings are those having a significant correlation at .05 with their factor (see Gorsuch, 1974). The fit for each set of comparisons is the average of the cosines of the angles of the corresponding pairs of items (see Kaiser et al, 1970). One factor was very stable across all of the

---

Insert Table 1 about here

---

interdivisional comparisons, with similarity coefficients of .82, .90, and .93. This factor is defined by "clearly indicates what material tests will cover," "clearly defines student responsibilities in the course," and "gives adequate information during the course regarding student progress through quizzes, tests, or other feedback." "Definition of Student Responsibility" seems a reasonably accurate name for this factor. One factor was stable across Humanities and Social Sciences, but not across the other comparisons. This appeared to be a "Broadmindedness" factor: "presents or allows various points of view," "welcomes criticisms from students," "invites criticisms of his own ideas," and "encourages class discussion." A rather difficult to name "Empathy/Clarity/Stimulation" factor was common to Humanities and Social Science: "Is concerned about the effectiveness of his teaching," "is genuinely interested in students," "welcomes questions from students," "is well informed on the materials presented," "clearly interprets abstract ideas and theories," "attempts to stimulate creative abilities," and "is enthusiastic about his subject." Finally, a weak "Course Coordination" factor seemed common to Social Science and Science/Mathematics; the two factors shared no salient items but were similar in overall pattern. The defining items were "keeps the course moving rapidly enough for the material," "makes it clear how each topic fits into the course," and "demands a reasonable amount of work."

According to the Kaiser statistic, then, only one factor -- "Definition of Student Responsibility" -- is stable across all interdivisional comparisons, although three other factors are common to one or another pairing of divisions. Visual inspection of Table 1 suggests that these partially generalizable factors (e.g., "Broadmindedness") do appear in other comparisons (e.g., in Science/Math and Social Science) and that additional factors seem sometimes to emerge (e.g., a "Stimulation" factor for Humanities and Social Sciences), but in all these cases the nuclear items are sometimes related to some items, other times to others, and generally quite difficult to interpret.

Comparisons within academic divisions. The results for the intradivisional comparisons -- factor similarity coefficients, indication of salient loadings and fit -- are presented in Table 2.

Insert Table 2 about here

There seems to be a generally greater stability of factor structures within divisions than between, in that a total of 11 intradivisional factor pairs met the .70 similarity criterion, compared to 6 pairs between divisions. At the same time, there is considerable variability in the intradivisional results. All 5 Humanities pairs met the criterion (range /.75/ to /.92/); five Social Science pairs were also very similar, but with a narrower range (/.71/ to /.78/); only 1 Science/Math pair had a coefficient of .70 or greater (/.56/ to /.75/). Fit also tended to be slightly higher within than across divisions.

The factor that was common across all divisions -- "Definition of Student Responsibility" -- seems clearly defined within each of the divisions too, although it failed to meet the similarity criterion in Science/Mathematics. An "Empathy/Clarity" or "Presence" factor was replicable for Humanities, as were ones describing "Broadmindedness" and "Stimulation." A fifth Humanities factor is difficult to interpret; it seems to be a second version of instructor presence. The "Empathy/Clarity" factor from the Humanities appeared in but was not replicable for Science/Math, and was quite different in Social Science, where a factor portraying clarity appeared in its own right. Conversely, a factor with items describing broadmindedness, empathy, and stimulation was replicable in Social Science, while one including broadmindedness and stimulation, but not empathy, was stable within Science/Mathematics. In short, the factors are for the most part specific to each academic division and even within divisions are not very generalizable.

DISCUSSION

Some comment about the numbers of instructors used in this study would be appropriate. A rule of thumb is not to factor analyze when the number of cases (instructors) is not five or ten times the number of variables (items), and especially not to factor analyze singular matrices (where the number of variables exceeds the number of cases). However, Rummel (1970, p. 220) points out that factor analysis of singular matrices allows descriptions of data variability, even though inference from sample results to universal factors is limited. Having more variables than cases imposes a necessary dependence on the interrelationships that can bias the inferences that could otherwise be drawn. The present study analyzes 26 variables for 15 and 25 cases, which would allow up to 15 (or 25) independent factors to emerge and that would certainly allow the major patterns of relationships to appear. Further support for this factor analysis of singular matrices comes from a computer-simulation study (Bejar & Doyle, in preparation) in which 25 variables were factor analyzed for 26, 20, 15, and 10 cases. Kaiser's factor-comparison procedure (1970) found a 5-factor solution (off-diagonal initial communalities, varimax rotation) recoverable even for the 10 cases. But in any case a replication of the present study would help determine its generalizability.

The question arises of why the present study found differential and limited factor generalizability while previous investigations (e.g., Isaacson et. al., 1964; Finkbeiner et al, 1973) found consistent and very high factor similarity. Some elaboration of the differences among within-class, between-instructors, and total covariance matrices may help resolve the apparent divergence of results, since the earlier studies analyzed total-variance matrices while the present one analyzed only the between-instructors matrix.

The sum of the variance-covariance matrices computed on each class weighted by their respective degrees of freedom (number of students in each minus 1) constitues the pooled within class sum of squares and cross products matrix (W). The variance-covariance matrix using class means (adjusting for class size) weighted by its degrees of freedom (number of classes minus 1) constitutes the between-instructor sum of square and cross products matrix (B). The sum of these two matrices equals the total sum of squares and cross products matrix, i.e.,

$$T = B + W$$

By appropriate rescaling, each of these 3 matrices can be converted to a correlation matrix and factor analyzed. Thus a factor analysis based on T reflects both within-class and between-instructor covariation. Whether W or B is more similar to T cannot be predicted beforehand. However if a) the degrees of freedom for the within-class source are large in relation to the between-instructor source and/or b) the within-class covariation is larger than the between-instructor, then T is more similar to W than to B. To the extent that these two conditions were fulfilled in the Isaacson and Finkbeiner studies -- and at least condition a) was -- the similarities found were from the within-class source. In the present study, only the between-instructor source was analyzed. Hence the apparently discrepant findings are the result of analyzing different sources of covariation.

The principal difference between within-class and between-instructors covariation lies in the computational treatment of student individual differences. Within classes, students' deviations from class means are treated as true variance. Between instructors, student differences are treated as error and are 'averaged out' by computation of the class means. Between-instructors data, then, more nearly describe instructor individual differences, while within-class data describe student differences. Thus halo effect and similar rater tendencies are more likely to be diminished in the between-instructors matrix, and so that matrix would seem to be the preferred one for many studies.

But cancelling out student differences by computing means removes not only some response tendencies but reliable statements of differential student/instructor interactions as well. So the portrayal of instructor individual differences by between-instructors data is accomplished at the cost of information about these interactions. The extent of this loss depends on the reliability of individual students' ratings and on the homogeneity of the class; the more reliable the ratings and the more heterogeneous the class, the greater is the loss of information about differential effectiveness. More research attention needs to be given to within-class data in general, but particularly to ways of increasing the reliability of individual student's ratings and to identifying patterns of differential student/instructor relationships. Similarly, the between-instructors matrix may provide a fruitful area of study, especially for the identification of effective instructional practices and for the validation of student ratings. It is suggested that the total variance matrix, because it confounds the between and within components of variance, be banished forever from the literature.

Table I

Interdivisional factor comparisons

Item	Sci & Math vs. Humanities[1]					Humanities vs. Soc Sci[2]					Sci & Math vs. Soc Sci[3]				
	(.66)	(.67)	(.87)	(.58)	(.63)	(.65)	(.79)	( )	( )	(.93)	(.59)	(.62)	(.90)	(.66)	(.79)
Is concerned about the effectiveness of his teaching.	a		b				b	a	b		a		b		
Is genuinely interested in student.	a			a			c	a	b		a	c		b	
Is well informed on the material presented.	a			a		b		a			a			b	
Clearly indicates what material tests will cover.			a							a			a		
Stimulates curiosity about the subject matter.	c	a			c	a		c			c	b			
Has an interesting style of presentation.	a	a	b			a		c			a	b	b		
Organizes his lectures well.	b	c		c		a		a			a			c	
Clearly interprets abstract ideas and theories.	a	c				a		a			a				
Attempts to stimulate creative abilities.	a	b				b		a			a				

Table 1 (continued)

Item	Sci & Math vs. Humanities[1]					Humanities vs. Soc Sci[2]					Sci & Math vs. Soc Sci[3]				
	(.66)	(.67)	(.87)	(.58)	(.63)	(.65)	(.79)	( )	( )	(.93)	(.59)	(.62)	(.90)	(.66)	(.79)
Keeps the course moving rapidly enough for the material.		c			b	b								c	b
Makes good use of examples and illustrations.	c			a		a		b	b					a	
Relates the material of this course with other areas of knowledge.		a				a	b	b				a			
Presents or allows various points of view.	b	a	c				a				b	a			
Discusses recent developments in the field.		a			c	a						b		c	
Is aware when students are having difficulty in understanding a topic.	c		a		c		c		c				a		
Makes it clear how each topic fits into the course.	a			c		c		b	c		a	b			c
Gives explanations which are clear and to the point.	a			c	c	a			b		b			c	
Welcomes questions from students.	a		c				a	a	b		b	c	c		
Is available for individual help.	a						c	b	b		b	c			

Table 1 (continued)

Item	Sci & Math vs. Humanities[1]					Humanities vs. Soc Sci[2]					Sci & Math vs. Soc Sci[3]				
	(.66)	(.67)	(.87)	(.58)	(.63)	(.65)	(.79)	( )	( )	(.93)	(.59)	(.62)	(.90)	(.66)	(.79)
Clearly defines student responsibilities in the course.			a	c		b		b		a			a		
Demands a reasonable amount of work.				b		a								b	c
Invites criticism of his own ideas.	a	a					a				b	a			
Is enthusiastic about his subject.	c	b		c				a			c	b			
Is humorous at appropriate times.		a				a						a			
Gives adequate information during the course regarding student progress through quizzes, tests, or other feedback.			a	b						a			a	b	
Encourages class discussion	c	a	c				a					b			
			Fit=.85					Fit=.85					Fit=.83		

[1] a=item was salient for both Sci & Math and Humanities; b=item was salient for Sci & Math; c=item was salient for Humanities.

[2] a=item was salient for both Humanities and Soc Sci; b=item was salient for Humanities; c=item was salient for Soc Sci.

[3] a=item was salient for both Sci & Math and Soc Sci; b=item was salient for Sci & Math; c=item was salient for Soc Sci.

Table 2

Intradivisional factor comparisons

Item	Science & Math[1]					Humanities[1]					Social Science[1]				
	( )	(.75)	(.63)	(.68)	( )	(.80)	(.92)	(.80)	(-.78)	(.75)	(.71)	(.76)	(.78)	(.73)	(.76)
Is concerned about the effectiveness of his teaching.	a		b				b	a	a		a	a	b		
Is genuinely interested in students.	a			b				a	b		c	a			
Is well informed on the material presented.	a			a		b		b	c		c	a			
Clearly indicates what material tests will cover.			a							a	c				
Stimulates curiosity about the subject matter.	c	a				a					a	a			
Has an interesting style of presentation.	b	b	b			a			c		a	c			
Organizes his lectures well.	b			c		a		a			a				
Clearly interprets abstract ideas and theories.	a			c		a	c	a			b	a			
Attempts to stimulate creative abilities.	a	a				a		b	c		a	a			

Table 2 (continued)

Item	Science & Math[1]					Humanities[1]					Social Science[1]				
	( )	(.75)	(.63)	(.68)	( )	(.80)	(.92)	(.80)	(.78)	(.75)	(.71)	(.76)	(.78)	(.73)	(.76)
Keeps the course moving rapidly enough for the material.					a	a		c			a			b	
Makes good use of examples and illustrations.				a		b	c	b	b						
Relates the material of this course with other areas of knowledge.		a				a	a	b				a		b	
Presents or allows various points of view.	b	a					a					a			
Discusses recent developments in the field.		a				b		c						b	
Is aware when students are having difficulty in understanding a topic.		c	a					c	a		c	a			
Makes it clear how each topic fits into the course.	b	a	c			c	c	a		c	a				
Gives explanations which are clear and to the point.	a			c		a			a	c	a	c			
Welcomes questions from students.	b						a	a	a			a			
Is available for individual help.	a		c					a	b			a			

Table 2 (continued)

Item	Science & Math[1]					Humanities[1]					Social Science[1]				
	( )	(.75)	(.63)	(.68)	( )	(.80)	(.92)	(.80)	(-.78)	(.75)	(.71)	(.76)	(.[illegible]3)	(.73)	(.76)
Clearly defines student responsibilities in the course.			a	c		b		b		a	c		b		
Demands a reasonable amount of work.	c		c	b		a									b
Invites criticism of his own ideas.	c	b	c				a					a			
Is enthusiastic about his subject.	c	b						a			c	b		c	
Is humorous at appropriate times.		a			c	b	c				a	b		a	
Gives adequate information during the course regarding student progress through quizzes, tests or other feedback.			a	a						a			a		
Encourages class discussion.		a					a					a			
		Fit=.88						Fit=.83					Fit=.88		

[1] a=the item was salient for both original and repetition samples; b=the item was salient only in the original sample; c=the item was salient only in the repetition sample.

REFERENCES

Bendig, A.W. A factor analysis of student ratings of psychology instructors on the Purdue Scale. Journal of Educational Psychology, 1954, 45, 385-393.

Bendig, A.W. An inverted factor analysis study of student-rated introductory psychology instructors. Journal of Experimental Education, 1953, 21, 333-336.

Cattel, R.B. Validity and Reliability: A Proposed more basic set of concepts. Journal of Educational Psychology, 1964, 55, 1-22.

Coffman, W.E. Determining students' concepts of effective teaching from their ratings of instructor. Journal of Educational Psychology, 1954, 45, 277-286.

Crannel, C.W. A preliminary attempt to identify the factors in student-teacher evaluation. Journal of Psychology, 1953, 36, 417-422.

Creager, J.A. A multiple factor analysis of the Purdue Rating Scale for Instructors. Unpublished Master's thesis, Purdue University, 1950.

Cronback, L.J.; Gleser, G.C.; Nanda, H. & Rajaratman, N. The dependability of behavioral measurement: Theory of generalizability for scores and profile. New York: John Wiley & SOns, 1972.

Deshpande, A.S.; Webb, S.C.; & Marks, E. Student perceptions of Engineering instructor behaviors and their relationship to the evaluation of instructor and courses. American Educational Research Journal, 1970, 289-305.

Doyle, K.O. & Liu, H. Factor Analysis of "Student Evaluation of Instruction, Part I," Measurement Services Center Studies, University of Minnesota, September 1972.

Finkbeiner, C.T.; Lathrop, J.S. & Schuerger, J.M. Course and instructor evaluation: Some dimensions of a questionnaire. Journal of Educational Psychology, 1973, 64, 159-163.

Gibb, C.A. Classroom behavior of the college teacher. Educational and Psychological Measurement, 1955, 15, 254-263.

Corsuch, R.L. Factor Analysis. Philadelphia: W.B. Saunders Company, 1974.

Guilford, V.P. Psychometric Methods, New York: McGraw-Hill, 1954.

Isaacson, R.L.; McKeachie, W.J.; Milholland, J.E.; Lin, Y.G.; Hoffeller, M.; Baerwaldt, J.W.; & Zinn, K.L. Dimensions of student evaluation of teachings. Journal of Educational Psychology, 1964, 55(6), 344-351.

Rummel, R.J. Applied factor Analysis. Evanston: Northwestern University Press, 1970.

Schneewind, K.A. & Cattell, R.B. Zum Problem der Faktoridentifikation: Verteilungen und Vertrauensin tervalle von Kongruenz Koeffizienten tur Oersonlich keitsfaktoren in Bereich ogjectiveanglytischer Tests. Psychologische Beitrage (Sonder druck), 1971, 12, 214-226.

Smalzreid, N.T. and Remmer, H.H. A factor analysis of the Purdue Rating Scale for Instructors. Journal of Educational Psychology, 1943, 34, 363-367.

Tucker,

Veldman, D.J. Fortran Programming for the behavioral Sciences. Chicago: Holt, Rinehart & Winston, 1967.

Wherry, R.J. Control bias in ratings. Department of the Army, The Adjutant General's Office, Personnel Research and Procedures Division, Personnel Research Branch, 1952, PRS Reports, 914, 915, 919, 920, 921.

FOOTNOTES

1. Request for reprints should be sent to Isaac I. Bejar, Measurement Services Center, 9 Clarence Avenue SE, Minneapolis, Minnesota 55414.
2. The authors are indebted to Dr. Susan Whitely for comments on an earlier version of this paper.
3. The authors are indebted to Dr. H.F. Kaiser for supplying a listing of the program for doing the factor comparisons. Veldman (1967) also lists a similar program (RELATE) although it is just for the case when two orthogonal factor matrices are compared.
4. The available computer programs cannot compute R as the estimate of communality when the correlation matrix is singular because the matrix cannot be inverted. However, Tucker and (1973) have provided a more general procedure which can be used with singular correlation matrices.